Model selection in random effects models for directed graphs using approximated Bayes factors

Zijlstra, B.J.H.; van Duijn, M.A.J.; Snijders, T.A.B.

Published in:
Statistica Neerlandica

DOI:
10.1111/j.1467-9574.2005.00283.x

Statistica Neerlandica (2005) Vol. 59, nr. 1, pp. 107–118

Model selection in random effects models for directed graphs using approximated Bayes factors

Bonne J. H. Zijlstra*, Marijtje A. J. van Duijn and Tom A. B. Snijders

Department of Sociology/Statistics and Measurement Theory, Heijmans Institute/ICS, University of Groningen, Grote Rozenstraat 31, 9712 TG, Groningen, the Netherlands

With the development of an MCMC algorithm, Bayesian model selection for the p_2 model for directed graphs has become possible. This paper presents an empirical exploration in using approximate Bayes factors for model selection. For a social network of Dutch secondary school pupils from different ethnic backgrounds it is investigated whether pupils report that they receive more emotional support from within their own ethnic group. Approximated Bayes factors seem to work, but considerable margins of error have to be reckoned with.

Key Words and Phrases: p_2 model, social network analysis, random effects, MCMC estimation.

1 Introduction

This paper investigates model selection using approximated Bayes factors applied to the p_2 model for directed graphs (VAN DUIJN, SNIJDERS and ZIJLSTRA, 2004). Bayes factors offer several advantages in model selection. Unlike in frequentist testing procedures, evidence in favor of a null hypothesis can be found. Non-nested models can be compared, and model uncertainty is taken into account when comparing multiple models.

With the development of MCMC algorithms (ZIJLSTRA, VAN DUIJN and SNIJDERS, in preparation) the p_2 model can be estimated adequately, this was previously not possible. From these algorithms, a sample from the posterior distributions of model parameters is obtained. Based on a sample from the posterior, reasonably straightforward methods are available to approximate Bayes factors like the harmonic mean of the likelihood and the BIC (see, e.g., KASS and RAFTERY (1995)). The harmonic mean, however, is known to give unstable estimates and the BIC may favor complex models over more parsimonious ones (BERGER and PERICCHI, 1996).

*B.J.H.Zijlstra@ppsw.rug.nl

In this paper, an empirical study is made of an approximation of Bayes factors using a proposal by NEWTON and RAFTERY (1994).

Model selection is applied to an empirical example of networks of high school pupils (BAERVELDT *et al.*, 2004). In the Dutch Social Behavior Study (Baerveldt and SNIJDERS, 1994), social network data were collected from 16–18 year old pupils belonging to the same year group. All high schools were so-called MAVO schools, which educate children of medium intellectual ability. In this study, one of the questions asked was: 'Which pupils help you when you are depressed, for example, after the end of a love affair or in a conflict with other people?' This type of question is often used in social network analysis to observe, for all ordered pairs (i, j) of individuals in a given group, whether or not i receives emotional support from j. Thus, binary networks of reported received emotional support were obtained, which can be analyzed by the p_2 model.

Here we use two networks from the Dutch Social Behavior Study with pupils from different ethnic backgrounds. One network is used as a calibration sample for the analysis of a second network. The two samples resemble each other with respect to ethnic and gender composition. However, the calibration sample with 62 pupils is larger than the analysis sample, containing 39 pupils. One of the research questions of the Dutch Social Behavior Study was whether reported emotional support is more prevalent among pupils from the same ethnic background. Here, we will consider this our main question, taking into account that support relations have been found to be more prevalent among pupils of the same gender (see, e.g., BAERVELDT *et al.*, 2004).

2 The p_2 model

The p_2 model is a model for the analysis of directed graphs that has been developed in the context of social network analysis (VAN DUIJN *et al.*, 2004). The directed graphs represent sent and received relationships. Nodes represent actors. The model assumes dependence between relations if the same actor is involved as a sender or as a receiver of the relations.

The unit of analysis for the p_2 model is a dyad: the pair of ties between two actors. Let $(Y_{ij} = y_{ij}, Y_{ji} = y_{ji})$ be the dyad of actors i and j, where Y_{ij} represents the tie indicator variable from actor i to actor j with binary outcome y_{ij}, and Y_{ji} the tie indicator variable from actor j to actor i with binary outcome y_{ji}. Each dyad has four possible outcomes:

$$(y_{ij}, y_{ji}) \in \{(1,0), (0,1), (1,1), (0,0)\}.$$

The p_2 model is an elaboration of the p_1 model (HOLLAND and LEINHARDT, 1981). This is a multinomial model for the four dyadic outcomes, where the log-odds of a relation from i to j depends linearly on sender i and receiver j, and vice versa. The log-odds of a mutual (1, 1) dyad is augmented by an interaction effect, representing reciprocity. Thus, the probabilities of the four outcomes are modeled as

$$P(Y_{ij}=y_{ij}, Y_{ji}=y_{ji})$$
$$=\frac{\exp\{y_{ij}(\mu+\alpha_i+\beta_j)+y_{ji}(\mu+\alpha_j+\beta_i)+y_{ij}y_{ji}\rho\}}{1+\exp\{\mu+\alpha_i+\beta_j\}+\exp\{\mu+\alpha_j+\beta_i\}+\exp\{\mu+\mu+\alpha_i+\beta_j+\alpha_j+\beta_i+\rho\}},$$
$$y_{ij}, y_{ji} \in \{0,1\}, \quad i,j=1,\ldots,n, \quad i \neq j. \tag{1}$$

Vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ contain actor-specific sender and receiver parameters, respectively. The parameter μ is called the density parameter. It represents the log-odds of a tie in the case of zero sender and receiver effects. Parameter ρ is the reciprocity parameter.

In the p_2 model, these parameters are further modeled to include covariates. The sender and receiver effects are no longer fixed parameters, but stochastic variables with a joint distribution, representing the dependence between relations from and to the same actor. The density and reciprocity parameters μ and ρ have subscripts i and j to indicate that these are dyad-specific. The sender, receiver, density and reciprocity effects are regressed on covariates:

$$\alpha_i = \mathbf{X}_{1i}\gamma_1 + A_i, \quad \beta_i = \mathbf{X}_{2i}\gamma_2 + B_i, \quad \mu_{ij} = \mu + \mathbf{Z}_{1ij}\boldsymbol{\delta}_1, \quad \rho_{ij} = \rho + \mathbf{Z}_{2ij}\boldsymbol{\delta}_2, \tag{2}$$

where A_i and B_i are random variables following a bivariate normal distribution with $E(A_i) = E(B_i) = 0$, variances σ_A^2 and σ_B^2, and covariance σ_{AB}. A_i and B_i will be called the random effects. $\mathbf{X_1}$ and $\mathbf{X_2}$ are matrices with actor-specific covariates and $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are matrices with dyad-specific covariates. Vectors γ_1 and γ_2 contain regression parameters for the sender and receiver effects, respectively. Vectors $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ contain regression parameters for the density and reciprocity effects, respectively.

3 Estimation of the p_2 model

The p_2 model is a generalized linear model with crossed-nested random effects. Previously, we used Iterative Generalized Least Squares (IGLS) algorithms for the estimation of the p_2 model (VAN DUIJN *et al.*, 2004). IGLS algorithms for generalized linear models (MCCULLAGH and NELDER, 1989) with random effects (see e.g. GOLDSTEIN, 1995), however, have been shown to give biased estimates (RODRÍGUEZ and GOLDMAN, 1995). Moreover, no accurate likelihood measures are provided.

Here we apply an MCMC algorithm that estimates the p_2 model well (ZIJLSTRA *et al.*, in preparation). It extends BROWNE's (1998) algorithm for logistic multilevel models, to multinomial cross-nested two-level models. Below, we will give a short description of the algorithm.

For the p_2 model we define three parameter sets: $\mathbf{C}$, the random effects, $\boldsymbol{\Sigma}$, the covariance matrix of the random effects, and $\boldsymbol{\theta}$, the fixed model parameters. $\mathbf{C}$ contains the random effects for all n actors, with the pair of random effects for a single actor i,

$$\mathbf{C}_i^{\mathrm{T}} = (A_i, B_i).$$

The covariance matrix of the random effects, $\boldsymbol{\Sigma}$, is

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}.$$

The vector $\boldsymbol{\theta}$ contains the fixed model parameters

$$\boldsymbol{\theta}^{\mathrm{T}} = (\mu, \rho, \gamma_1{}^{\mathrm{T}}, \gamma_2{}^{\mathrm{T}}, \boldsymbol{\delta}_1{}^{\mathrm{T}}, \boldsymbol{\delta}_2{}^{\mathrm{T}}).$$

Bayesian estimation requires specification of prior distributions for the model parameters. For $\boldsymbol{\Sigma}^{-1}$ we assume an inverse Wishart distribution with v_σ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}_\sigma$ and for $\boldsymbol{\theta}$ we assume a normal prior distribution.

The posterior distribution is approximated by an MCMC algorithm that is comparable to a Gibbs sampler (see, e.g., Chib and GREENBERG, 1995) where subsequent draws from the conditional distributions for all parameter sets are taken. The conditional distribution for the random effects is

$$P(\mathbf{C}|\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\Sigma}) \propto \prod_{i<j}^{n} \{f_1(Y_{ij}, Y_{ji}|\mathbf{C}, \boldsymbol{\theta})\} f_2(\mathbf{C}|\boldsymbol{\Sigma}),$$

where f_1 is (1) with substitutions as in (2) and f_2 is the normal density for all random effects. For the conditional distribution of $\boldsymbol{\Sigma}$ we draw from the conditional distribution of $\boldsymbol{\Sigma}^{-1}$ and invert this draw afterwards (see, e.g., GILKS, RICHARDSON and SPIEGELHALTER, 1996):

$$(\boldsymbol{\Sigma}^{-1}|\mathbf{Y}, \mathbf{C}, \boldsymbol{\theta}) \sim \text{ Wishart}\left(v_\sigma + n, \left(\sum_{i=1}^{n} \mathbf{C}_\mathrm{i} \mathbf{C}_\mathrm{i}^T + \boldsymbol{\Sigma}_\sigma^{-1} \right)^{-1} \right). \tag{3}$$

The fixed model parameters, $\boldsymbol{\theta}$, have conditional distribution

$$P(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{C}, \boldsymbol{\Sigma}) \propto \prod_{i<j}^{n} \{f_1(Y_{ij}, Y_{ji}|\mathbf{C}, \boldsymbol{\theta})\} f_3(\boldsymbol{\theta}),$$

where f_3 is the normal prior density of theta.

The conditional distributions of the random effects, $\mathbf{C}$, and the fixed model parameters, $\boldsymbol{\theta}$, cannot be simulated directly. Therefore, like BROWNE (1998) did in logistic random effects models, we approximate sampling from these distributions using a Metropolis algorithm with a random walk proposal distribution. This is a normal distribution with zero means and some covariance matrix. The covariance matrix is adapted to result in an optimal acceptance ratio for the Metropolis steps. We took this ratio to be 1/3 (see GELMAN, ROBERTS, and GILKS 1995).

4 Estimating the calibration sample

The calibration sample will be used to obtain prior distributions for the analysis sample. It was chosen such that its ethnic composition resembles that of the analysis sample. The network consists of 62 pupils with 37 boys, and 18 pupils with a Dutch ethnic background, 6 Moroccan, 14 Turkish and 12 Surinamese.

For the prior distribution of $\mathbf{\Sigma}^{-1}$ of the calibration sample we took $v_\sigma = 3$ and $\mathbf{\Sigma}_\sigma = \mathbf{I}$, knowing that an identity matrix for the covariance of the random effects is not far from what is commonly observed. The prior distribution of $\boldsymbol{\theta}$ for the calibration sample is normal with zero means and a diagonal covariance matrix with variances for μ and ρ equal to 100. The variances of the regression parameters γ_1, γ_2, $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are chosen as 100 divided by the variance of the corresponding covariate. Thus, the variance of a parameter of a 'standardized' covariate would also be equal to 100. Since parameters in $\boldsymbol{\theta}$ are on a logscale, a variance of 100 here implies that 33% of the observations are larger than 10 in absolute value, which is a very large value. Therefore, the prior for $\boldsymbol{\theta}$ represents the quite vague prior information about the likely values of this vector of parameters.

Table 1 shows the results from the MCMC algorithm for the calibration sample. Gender and ethnic background are actor-specific covariates, which can be transformed into dichotomous dyadic covariates indicating whether or not two pupils have the same gender or ethnic background. For the sender and receiver effects, gender is a dummy variable, where boys have code one and girls code zero. The estimates for the calibration sample in Table 1 are for the full model, which is the most elaborate of the models under consideration. Results are based on 30 000 iterations following a burn-in sample of 10 000.

From the estimates below, girls are more often reported as giving emotional support. Also, more emotional support is reported between pupils of the same gender and with Surinamese parents.

Table 1. Parameter estimates for the calibration sample.

			Calibration school		Analysis sample	
			Posterior		Prior distribution	
Effect	Covariate	Parameter	Mean	(S.E.)	Mean	(S.D.)
Sender	Gender	γ_{11}	0.18	(0.50)	0.18	(1.5)
Receiver	Gender	γ_{21}	−1.26	(0.49)	−1.26	(1.5)
Density		μ	−3.32	(0.36)	−3.32	(2.5)
	Gender	δ_{11}	1.29	(0.41)	0.5	(1.5)
	Dutch	δ_{12}	0.24	(0.33)	0.5	(1.5)
	Moroccan	δ_{13}	0.78	(0.50)	0.5	(1.5)
	Turkish	δ_{14}	0.65	(0.34)	0.5	(1.5)
	Surinamese	δ_{15}	1.42	(0.44)	0.5	(1.5)
Reciprocity		ρ	4.20	(0.65)	4.20	(2.5)
	Gender	δ_{21}	−1.25	(0.85)	0	(2)
	Dutch	δ_{22}	1.10	(1.00)	0	(2)
	Moroccan	δ_{23}	−1.06	(1.20)	0	(2)
	Turkish	δ_{24}	0.16	(0.88)	0	(2)
	Surinamese	δ_{25}	−0.84	(1.13)	0	(2)
Sender variance		σ_A^2	0.69	(0.28)	1	
Receiver variance		σ_B^2	0.44	(0.23)	1	
Sender receiver covariance		σ_{AB}	−0.09	(0.19)	0	

5 Bayes factors

Bayes factors (Kass and Raftery, 1995) provide evidence either in favor of or against a hypothesis. Unlike frequentist test procedures, the Bayes factor does not evaluate one model conditional on another, but can be used to compare a number of non-nested models.

Model A, M_A, can be compared with Model B,M_B, by taking the posterior odds of model A versus model B given the data $\mathbf{Y}$,

$$\frac{P(M_A|\mathbf{Y})}{P(M_B|\mathbf{Y})} = \frac{P(\mathbf{Y}|M_A)}{P(\mathbf{Y}|M_B)} \times \frac{P(M_A)}{P(M_B)},$$

in words, posterior odds = Bayes factor $\times$ prior odds. Thus, the Bayes factor for Model A versus Model B is defined as

$$B_{AB} = \frac{P(\mathbf{Y}|M_A)}{P(\mathbf{Y}|M_B)}.$$

Note that if the prior odds for the models are equal to one, the posterior odds equal the Bayes factor.

Different suggestions have been made for interpreting the values of the Bayes factors. Following Raftery's (1996) interpretation, positive support for Model A versus Model B is found if the natural log of the Bayes factor is between 1.1 and 3. A logarithm of the Bayes factor between 3 and 5 is interpreted as strong support. A larger Bayes factor indicates even stronger support.

6 Approximating Bayes factors

Calculating Bayes factors involves calculating the probability of the data given a model k, $P(\mathbf{Y}|M_k)$, which is called the marginal likelihood. It is obtained by integrating over the parameter space under model k. For the p_2 model this gives

$$P(\mathbf{Y}|M_k) = \int\int P(\mathbf{Y}|\boldsymbol{\theta}_k, \mathbf{C}_k, M_k)P(\boldsymbol{\theta}_k, \mathbf{C}_k|M_k)\mathrm{d}\mathbf{C}_k\,\mathrm{d}\boldsymbol{\theta}_k, \tag{4}$$

where $P(\boldsymbol{\theta}_k, \mathbf{C_k}|M_k)$ is the joint prior distribution for $\boldsymbol{\theta}$ and $\mathbf{C}$ marginalized over $\boldsymbol{\Sigma}$. For the p_2 model this integral cannot be computed analytically, but needs to be approximated.

At first sight a straightforward approximation to (4) appears be the Monte Carlo integral

$$\frac{1}{T}\sum_{t=1}^{\mathrm{T}} P(\mathbf{Y}|\boldsymbol{\theta}_k{}^{(t)}, \mathbf{C}_k^{(t)}, M_k), \tag{5}$$

with $\boldsymbol{\theta}_k{}^{(t)}$ the t^{th} draw of $\boldsymbol{\theta}_k$ from the posterior distribution available from the MCMC estimation. However, because draws from the posterior distribution of the parameters are all conditional on the data $\mathbf{Y}$, (5) is in fact the estimate of

$$\int\int P(\mathbf{Y}|\boldsymbol{\theta}_k, \mathbf{C}_k, M_k) P(\boldsymbol{\theta}_k, \mathbf{C}_k|M_k, \mathbf{Y}) \mathrm{d}\mathbf{C}_k \, \mathrm{d}\boldsymbol{\theta}_k,$$

which can be rewritten as the rather awkward integral

$$\int\int P(\mathbf{Y}|\boldsymbol{\theta}_k, \mathbf{C}_k, M_k)^2 \frac{P(\boldsymbol{\theta}_k, \mathbf{C}_k|M_k)}{P(\mathbf{Y}|M_k)} \mathrm{d}\mathbf{C}_k \, \mathrm{d}\boldsymbol{\theta}_k,$$

the *posterior mean* of the likelihood (AITKIN, 1991). Taking the *harmonic mean* of the likelihood

$$\widehat{Pr_1}(\mathbf{Y}|M_k) = \left(\frac{1}{T} \sum_{t=1}^{T} P(\mathbf{Y}|\boldsymbol{\theta}_k{}^{(t)}, \mathbf{C}_k^{(t)}, M_k)^{-1} \right)^{-1}$$

results in a correct estimator of (4). The harmonic mean converges almost surely to the correct distribution $P(\mathbf{Y}|M_k)$, but is found to be unstable because of its sensitivity to outliers; occasional parameters with a small likelihood. Put differently, the inversion applied in the harmonic mean estimator may give rise to a long-tailed distribution.

NEWTON and RAFTERY (1994) suggest a weighted estimator based on T values from the posterior and $\delta T/(1 - \delta)$ imaginary values draw from the prior distribution of the parameters in $\boldsymbol{\theta}$. The imaginary draws have a likelihood $P(\mathbf{Y}|\boldsymbol{\theta}_k, \mathbf{C}_k)$ equal to their expected value $P(\mathbf{Y}|M_k)$. This gives the estimator

$$\widehat{Pr_4}(\mathbf{Y}|M_k) = \frac{\dfrac{\delta T}{1-\delta} + \displaystyle\sum_{t=1}^{T} \frac{P(\mathbf{Y}|\boldsymbol{\theta}_k{}^{(t)}, \mathbf{C}_k^{(t)}, M_k)}{\delta \widehat{Pr_4}(\mathbf{Y}|M_k) + (1-\delta) P(\mathbf{Y}|\boldsymbol{\theta}_k{}^{(t)}, \mathbf{C}_k^{(t)}, M_k)}}{\dfrac{\delta T}{(1-\delta)\widehat{Pr_4}(\mathbf{Y}|M_k)} + \displaystyle\sum_{t=1}^{T} \frac{1}{\delta \widehat{Pr_4}(\mathbf{Y}|M_k) + (1-\delta) P(\mathbf{Y}|\boldsymbol{\theta}_k{}^{(t)}, \mathbf{C}_k^{(t)}, M_k)}}.$$

An iterative scheme is applied to obtain the estimate of the recursive $\widehat{Pr_4}(\mathbf{Y}|M_k)$ (we use the subscript 4 in accordance with NEWTON and RAFTERY (1994) and KASS and RAFTERY (1995)). As $\delta \downarrow 0, \widehat{Pr_4}(\mathbf{Y}|M_k)$ approaches the harmonic mean estimator, $\widehat{Pr_1}(\mathbf{Y}|M_k)$. For $\widehat{Pr_4}(\mathbf{Y}|M_k)$ it is important to find a value of δ for which the estimate does not display the high sensitivity to small values of the likelihood that are the problem of Pr_1.

Because Bayes factors in model selection are sensitive to the choice of the prior distribution, it is recommended that one use non-vague priors (see, e.g., BERGER and PERICCHI (1996)). Here we will use the calibration sample to obtain reasonable parameters for the prior distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}^{-1}$.

The prior distributions for the parameters in $\boldsymbol{\theta}$ are shown in Table 1. They are chosen as rounded versions of the posteriors obtained from the calibration sample, with higher standard deviations representing the possibility that the new year group (analysis sample) differs from the one studied before.

Table 2. Largest and smallest observed values of the natural logarithm of Pr_4 in ten replications of MCMC estimation of the Full Model for two different chain lengths.

Burn-in/Sample	Pr_4	$\delta = 0$	$\delta = 0.001$	$\delta = 0.01$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.5$
10 000/30 000	Largest	−253.57	−252.47	−249.98	−247.35	−245.66	−243.16	−238.48
	Smallest	−257.80	−256.74	−254.70	−252.11	−249.97	−247.20	−242.06
20 000/40 000	Largest	−253.80	−252.74	−251.21	−247.85	−245.90	−243.23	−238.90
	Smallest	−259.73	−257.84	−255.14	−251.52	−249.02	−246.38	−242.45

For μ, ρ, γ_1, and γ_2, the observed posterior means from the calibration school are taken. Parameters μ and ρ can be sensitive to sample size because typically in larger networks the probability of a relation decreases. Therefore, their standard errors are increased roughly by two. Standard errors of γ_1 and γ_2 are increased approximately by one compared with the calibration sample. The parameters in $\boldsymbol{\delta}_1$ all have prior means 0.5, roughly summarizing from the calibration sample that emotional support is more prevalent in dyads whose actors have the same gender or ethnic background. The posterior standard errors are increased by approximately 1. Prior means for parameters in $\boldsymbol{\delta}_2$ are chosen to be zero because of the undecisive estimates obtained from the calibration sample. Again, approximately 1 is added to the standard errors of the calibration sample.

For the prior distribution of $\boldsymbol{\Sigma}^{-1}$, we again took the identity matrix with three degrees of freedom. We did not take the observed covariance matrix from the calibration school, because the algorithm is sensitive to small values of $\boldsymbol{\Sigma}_\sigma^{-1}$; it appears to move more slowly through sampled values with small cross-products $\mathbf{C}_i\mathbf{C}_i^{\mathrm{T}}$ in (3).

6.1 Determining δ for the analysis sample

When Pr_4 is used for model selection, it is important to establish what would be a good value for δ. For this purpose the full model (see Table 3) for the analysis sample is replicated twenty times using different starting values. This gives a rough indication of the variation displayed by Pr_4 for different values of δ. Results are given in Table 2. The first ten replications are based on 30 000 iterations of the MCMC algorithm and a burn-in sample of 10 000. The last ten replications are based on 40 000 iterations and a burn-in sample of 20 000.

Most variation is observed for $\delta = 0$, for which Pr_4 coincides with the harmonic mean estimator Pr_1. For all values of δ larger than zero, the observed variation in Pr_4 is smaller. Note, however, that for all these values of δ a difference as large as 4 is observed. Here we will use $\delta = 0.01$ for which Newton and Raftery (1994) also mention that Pr_4 performs well.

7 Model selection for the analysis sample

The network used for model selection contains 39 pupils, 16 of which are boys. There are 6 pupils from a Dutch background, 5 from a Moroccan, 8 from a Turkish, and 11 from a Surinamese background.

Five models with decreasing complexity will be compared. The first model is the full model, taking into account all possible effects of gender (as sender covariate, receiver, density and reciprocity covariate), as well as the four categories for ethnic background as density and reciprocity covariates.

The second model is equal to the full model, except that it does not contain any reciprocity covariates. (The experience with the p_2 model so far is that dyadic covariate effects for reciprocity are seldom found.) The third model contains only sender, receiver and density effects for gender. The fourth model is equal to Model 3, with an additional density effect of ethnic background. (As for the other density covariates, the prior distribution for the regression parameter for this additional covariate is normal with a mean of 0.5 and a standard deviation of 1.5.) In contrast to Model 2, no distinction is made between the different ethnic groups. The fifth and final model is the empty model, without covariates. Tables 3 and 4 present parameter estimates of the p_2 model for the analysis sample.

From the parameter estimates of the full model in Tables 3 and 4, it appears that boys more often report received emotional support than girls, but girls are more often reported as giving emotional support. Furthermore, pupils report more received emotional support from within their own ethnic background, but within the same gender or ethnic background there is no evidence of a stronger tendency to mutually reported support relations.

To investigate if there is support for separate parameters for the different categories of ethnic background, Model 4 includes a single covariate that indicates whether pupils have the same ethnic background.

Table 3. Parameter estimates the Full Model and Model 2.

		Full Model		Model 2	
		Posterior		Posterior	
Effect	Covariate	Mean	(S.E.)	Mean	(S.E.)
Sender	Gender	0.91	(0.35)	1.07	(0.46)
Receiver	Gender	−0.80	(0.41)	−0.98	(0.41)
Density		−3.40	(0.37)	−3.22	(0.44)
	Gender	0.57	(0.36)	0.89	(0.26)
	Dutch	1.05	(0.46)	0.83	(0.30)
	Moroccan	0.25	(0.47)	0.30	(0.29)
	Turkish	0.82	(0.38)	0.72	(0.25)
	Surinamese	0.38	(0.34)	0.59	(0.24)
Reciprocity		4.70	(0.66)	4.40	(0.69)
	Gender	0.78	(0.82)		
	Dutch	−0.78	(1.02)		
	Moroccan	0.03	(0.92)		
	Turkish	−0.32	(0.91)		
	Surinamese	0.52	(0.78)		
Sender variance		0.65	(0.34)	0.78	(0.40)
Receiver variance		0.61	(0.31)	0.58	(0.36)
Sender receiver covariance		−0.42	(0.28)	−0.37	(0.37)
Pr_4 ($\delta = 0.01$)		−253.54		−254.07	

Table 4. Parameter estimates for Models 3 and 4 and the Empty Model.

		Model 3		Model 4		Empty Model	
		Posterior		Posterior		Posterior	
Effect	Covariate	Mean	(S.E.)	Mean	(S.E.)	Mean	(S.E.)
Sender	Gender	0.96	(0.38)	1.00	(0.42)		
Receiver	Gender	−0.83	(0.39)	−0.90	(0.39)		
Density		−4.01	(0.31)	−4.23	(0.32)	−4.29	(0.29)
	Gender	0.76	(0.23)	0.83	(0.23)		
	Same ethnic background			1.08	(0.21)		
Reciprocity		4.78	(0.59)	4.70	(0.59)	5.02	(0.58)
Sender variance		0.67	(0.33)	0.82	(0.36)	0.91	(0.42)
Receiver variance		0.59	(0.30)	0.58	(0.29)	0.80	(0.35)
Sender receiver covariance		−0.45	(0.28)	−0.51	(0.29)	−0.69	(0.34)
Pr_4 ($\delta = 0.01$)		−266.70		−250.05		−265.63	

Table 5. Natural logarithm of Bayes factors for high school data.

$M_A \backslash M_B$	Full model	Model 2	Model 3	Model 4	Empty Model
Full model		0.53	13.06	−3.49	12.09
Model 2	−0.53		12.63	−4.02	11.56
Model 3	−13.06	−12.63		−16.56	−1.04
Model 4	3.49	4.02	16.65		15.58
Empty model	−12.09	−11.56	1.04	−15.58	
Posterior probability	0.029	0.017	5.6E-8	0.954	1.6E-7

From the estimates of Pr_4 in Tables 3 and 4, Bayes factors can be computed. These are displayed in Table 5, where also the posterior model probabilities are given. These probabilities are calculated as

$$P(M_k|\mathbf{Y}) = \frac{P(\mathbf{Y}|M_k)P(M_k)}{\sum_{l=1}^{K} P(\mathbf{Y}|M_l)P(M_l)}.$$

That is, given K models and equal prior odds for these models, the posterior model probability is the marginal likelihood for one model, relative to the added marginal likelihoods of all models under consideration.

The results in Table 5 indicate that the full model and Models 2 and 4 are clearly preferred over Model 3 and the empty model. Comparing Model 4 with Model 2, there is indeed no evidence that the data support a model with differential effects on the density covariate for the different ethnic backgrounds. However, comparing the full model with Models 2 and 4, the maximal difference of Pr_4 between any of these models hardly exceeds 4, the random fluctuation observed in Table 2. Considering that the log of the Bayes factor is the difference between the log of two estimates of Pr_4, there is no conclusive evidence in favour of either the full model, Model 2 or Model 4. In contrast, there is clearly evidence against Model 3 and the empty model. A natural way to deal with competing models is to prefer the most parsimonious model. This makes Model 4 the preferred model over Model 2 and the full model.

In conclusion of the model selection of these high school data, there is evidence that boys more often than girls report having received emotional support, while emotional support is less often reported to come from boys. From pupils with the same gender and ethnic background, more received emotional support is reported. However, there is no evidence that this effect is different for different ethnic groups.

8 Concluding remarks

The Pr_4 estimator of Newton and Raftery (1994) proves to be helpful, although it combines some nice properties with some disturbing ones. In our view Pr_4 did a good job because it allowed us to show that reported emotional support is more likely from within the same ethnic background, but there is no evidence that this effect differs between the categories of ethnic background.

One disturbing property of Pr_4 is that it depends heavily on the value of δ. If δ equals zero, Pr_4 equals the harmonic mean estimator Pr_1 and if δ approaches 1, the Pr_4 approaches the posterior mean of the likelihood. The latter parameter can be used to compose a 'posterior Bayes factor' (Aitkin, 1991). The posterior mean of the likelihood is not an estimate of the marginal likelihood. In our example, Pr_4 with $\delta > 0$ displays slightly less variation compared with the harmonic mean estimator. However, it does put us on a sliding scale towards a posterior Bayes factor, which is undesirable.

Finally, it should be noted that from replicating model estimates for our example, for all $\delta > 0$, fluctuations of Pr_4 as large as 4 were observed. Such fluctuations in Pr_4 lead to approximate Bayes factors that are interpreted as strong evidence. Clearly, random fluctuations must be considered in evaluating Bayes factors and estimates of the size of these are needed. Also further research into approximate Bayes factors for the p_2 model that have smaller errors is required. One method that needs further attention is the approximation of the marginal likelihood as proposed by Chib and Jeliazkov (2001), possibly in combination with other estimation methods under development.

Acknowledgements

The authors would like to thank Chris Baerveldt for his permission to use the data. We would also like to thank the organizers and participants of the workshop on Bayesian Model Selection in Utrecht, July 2004, for their many useful suggestions.

References

Aitkin, M. (1991), Posterior Bayes factors, *Journal of the Royal Statistical Society* B **53**, 111–142.

Baerveldt, C., M. A. J. van Duijn, L. Vermeij and D. A. van Hemert (2004), Ethnic boundaries and personal choice. Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils' networks, *Social Networks* **26**, 55–74.

Baerveldt, C. and T. A. B. Snijders (1994), Influences on and from the segmentation of networks: hypotheses and tests, *Social Networks* **16**, 213–232.

Berger, J. O. and L. R. Pericchi (1996), The intrinsic Bayes factor for model selection and prediction, *Journal of the American Statistical Association* **91**, 109–122.

Browne, W. J. (1998) *Applying MCMC methods to multi-level models*, PhD thesis, University of Bath.

Chib, S. and I. Jeliazkov (2001), Marginal likelihood from the Metropolis-Hastings output, *Journal of the American Statistical Association* **96**, 270–281.

Chib, S. and E. Greenberg (1995), Understanding the Metropolis–Hastings algorithm, *The American Statistician* **49**, 327–335.

Gelman, A., G. O. Roberts and W. R. Gilks (1995), Efficient Metropolis jumping rules, in: J. M. Bernardo, J. O. Berger, A. P. David and A. F. M. Smith (eds), *Bayesian Statistics 5*, Oxford University Press, Oxford, 599–607.

Gilks, W. R., S. Richardson and D. J. Spiegelhalter (eds), (1996), *Markov chain Monte Carlo in practice*, Chapman & Hall, London.

Goldstein, H. (1995), *Applied multilevel analysis* (2nd edn), Edward Arnold, London.

Holland, P. W. and S. Leinhardt (1981), An exponential family of probability distributions for directed graphs, *Journal of the American Statistical Association* **77**, 33–50.

Kass, R. E. and A. E. Raftery (1995), Bayes factors, *Journal of the American Statistical Association* **90**, 773–795.

McCullagh, P. and J. A. Nelder (1989), *Generalized linear models*, Chapman & Hall, London.

Newton, M. A. and A. E. Raftery (1994), Approximate Bayesian inference with the weighted likelihood bootstrap, *Journal of the Royal Statistical Society* B **56**, 3–48.

Raftery, A. E. (1996), Approximate Bayes factors and accounting for model uncertainty in generalized linear models, *Biometrika* **83**, 251–266.

Rodríguez, G. and N. Goldman (1995), An assessment of estimation procedures for multilevel models with binary responses, *Journal of the Royal Statistical Society* A **158**, 73–89.

Van Duijn, M. A. J., T. A. B Snijders and B. J. H. Zijlstra (2004), p_2: a random effects model with covariates for directed graphs, *Statistica Neerlandica* **58**, 234–254.

Zijlstra, B. J. H., M. A. J. van Duijn and T. A. B. Snijders. MCMC estimation of the p_2 model for directed graphs with covariates, in preparation.

Received: June 2004. Revised: December 2004.